Dear Referee,

We are grateful for the referee's comments. We have prepared detailed responses to each comment and outlined the specific revisions that we will incorporate into the manuscript in the revised submission. The major updates include: **(i) a new intercomparison section comparing our reconstruction with GOBAI and ITO using the filtered GLODAPv2 dataset as an independent validation benchmark, together with new profile- and map-based comparisons against WOA23 (Figs. R1-R2; Table R1); (ii) a new figure showing the spatial distribution and temporal histograms of the 8,020 unique profiles in the filtered GLODAPv2 validation set to clarify its spatial-temporal representativeness (Fig. R3); (iii) a revision of the original dynamic weighting framework by replacing the previous hard switch with a spatially smoothed continuous weighting scheme, together with a comparison of the resulting weight fields and the DO fields (Fig. R4-R5); and (iv) a depth-specific assessment of deep-ocean accuracy through newly added comparisons with the DIVA-based Roach and Bindoff product (Figs. R6-R8).** A point-by-point reply follows below (referee comments in *italics*, our responses in regular type).

On behalf of all authors, sincerely,

Mingyu Han
Shanghai Jiao Tong University

Comment 1:

1. Lack of Intercomparison with Existing Data Products

The manuscript currently lacks a robust comparison with other widely used ocean DO data products. To establish the reliability of this new product, it is essential to contextualize its performance against existing datasets. I strongly recommend adding a comprehensive data intercomparison section to validate the BLENDR outputs. The authors should refer to and compare their results within the context of recent multi-product coordinated intercomparisons, such as the one presented by Ito et al. (2025). This will significantly enhance the credibility of your product.

Reference: Ito T, Garcia H E, Wang Z, et al. Assessing the observational uncertainties of dissolved oxygen climatology and seasonal cycle through a coordinated intercomparison project[J]. Global Biogeochemical Cycles, 2025, 39(11): e2025GB008751.

Response 1:

We thank the reviewer for this suggestion. In response, we added an intercomparison between our DO reconstruction and two recent products, GOBAI (Sharp et al., 2023) and ITO (Ito et al., 2024), using the filtered GLODAPv2 dataset (Olsen et al., 2016) as an independent validation benchmark. The results show that, at the global scale, our reconstruction achieves the lowest MAE and RMSE and the highest R^2 among the three products, indicating the best overall agreement with independent observations (Table R1). Within the GOBAI coverage, our reconstruction has a lower RMSE and higher R^2 than GOBAI, while its MAE is slightly higher and its mean difference is closer to zero. Within the ITO coverage, our reconstruction has lower MAE and RMSE and higher R^2 than ITO, while its mean difference is farther from zero. These results support the reliability of our product.

Table R1. Performance comparison on the filtered GLODAPv2

Product	MAE	RMSE	R^2	ΔDO
Our reconstruction	10.316	18.212	0.967	-0.276
GOBAI on filtered GLODAPv2	**11.101**	19.875	0.956	-0.971
Our reconstruction in GOBAI coverage	11.236	**19.731**	**0.963**	**-0.399**
ITO on filtered GLODAPv2	13.415	22.958	0.951	**-0.123**
Our reconstruction in ITO coverage	**11.937**	**20.045**	**0.964**	-0.485

We also added a profile-based comparison. Figure R1a shows the global mean vertical profiles of dissolved oxygen from our reconstruction, ITO Oxygen (Ito et al., 2024), and WOA23 climatology (Garcia et al., 2024) over 1965 – 2022. Near the surface, the three profiles are very close. In the 800-1000 m depth range, our reconstruction is close to WOA23, while ITO Oxygen is

lower over part of this depth range. Below 1000 m, ITO Oxygen does not provide data, so the comparison is limited to our reconstruction and WOA23. Their profiles remain close through the deep ocean, indicating that our product gives a reasonable extension of dissolved oxygen fields below the depth range covered by ITO Oxygen. Figure R1b shows the global mean vertical profiles of our reconstruction, ITO Oxygen, and GOBAI over 2004–2020 for the upper 2000 m. The three products show a similar overall vertical structure, with the largest differences appearing in the 500-1000 m depth range. In this depth range, our reconstruction is generally higher than both ITO Oxygen and GOBAI, while the three profiles are closer near the surface. This comparison shows that our product reproduces the large-scale vertical pattern seen in existing datasets, while also showing differences in intermediate waters.

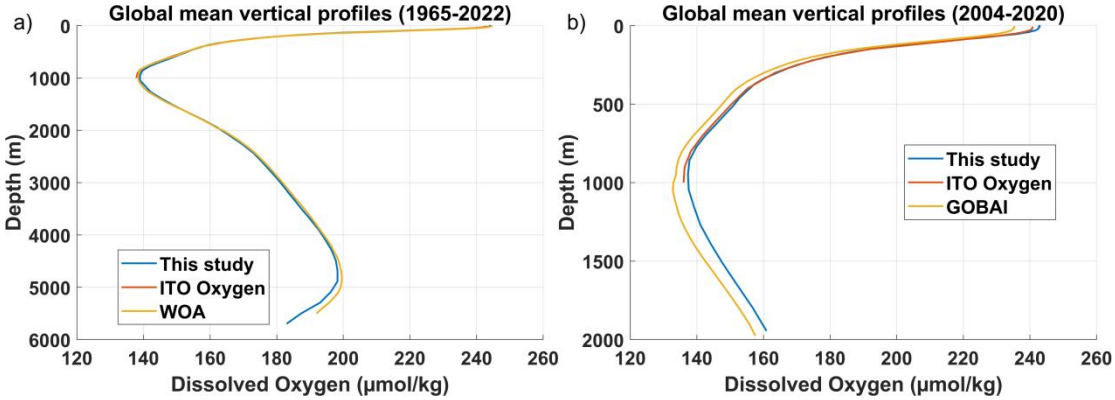

Figure R1. Global mean vertical profiles of dissolved oxygen from different products. (a) Profiles for this study, ITO Oxygen, and WOA23 over 1965–2022, shown from the surface to 5902 m. (b) Profiles for this study, ITO Oxygen, and GOBAI over 2004–2020, shown from the surface to 2000 m.

We further added a spatial comparison with the WOA23 climatology at several representative depths (Figure R2). Our reconstruction is closer to WOA23, particularly in the surface layer, and shows smaller differences in many low- and mid-latitude regions. At the surface layer around 10 m depth, our reconstruction shows small differences, generally within ±2 μmol kg^{-1}, except in some high-latitude regions. In comparison, ITO Oxygen exhibits broader regions of red, corresponding to negative differences of about 4–8 μmol kg^{-1} in the subtropical gyres, and more pronounced blue regions, corresponding to positive differences of about 6–10 μmol kg^{-1} under the Antarctic Circumpolar Current. At 30 m, the differences in our reconstruction remain small in the mid-latitude regions, with larger variability near boundary currents. In contrast, ITO Oxygen again shows larger negative differences in the subtropics and positive differences in the southern high latitudes. These results indicate that our reconstruction is generally closer to WOA23 in the surface ocean. At around 200 m, both our reconstruction and ITO Oxygen show larger departures from the WOA23 reference, reaching about ±10 μmol kg^{-1} in the tropical and subtropical regions. At around 700 m, our reconstruction and WOA23 remain within about ±8 μmol kg^{-1} over large parts of the Atlantic and Pacific basins, indicating good agreement at mid-depths. These spatial maps complement the statistical comparisons by showing that our product remains close to a widely used climatological reference across multiple depth levels.

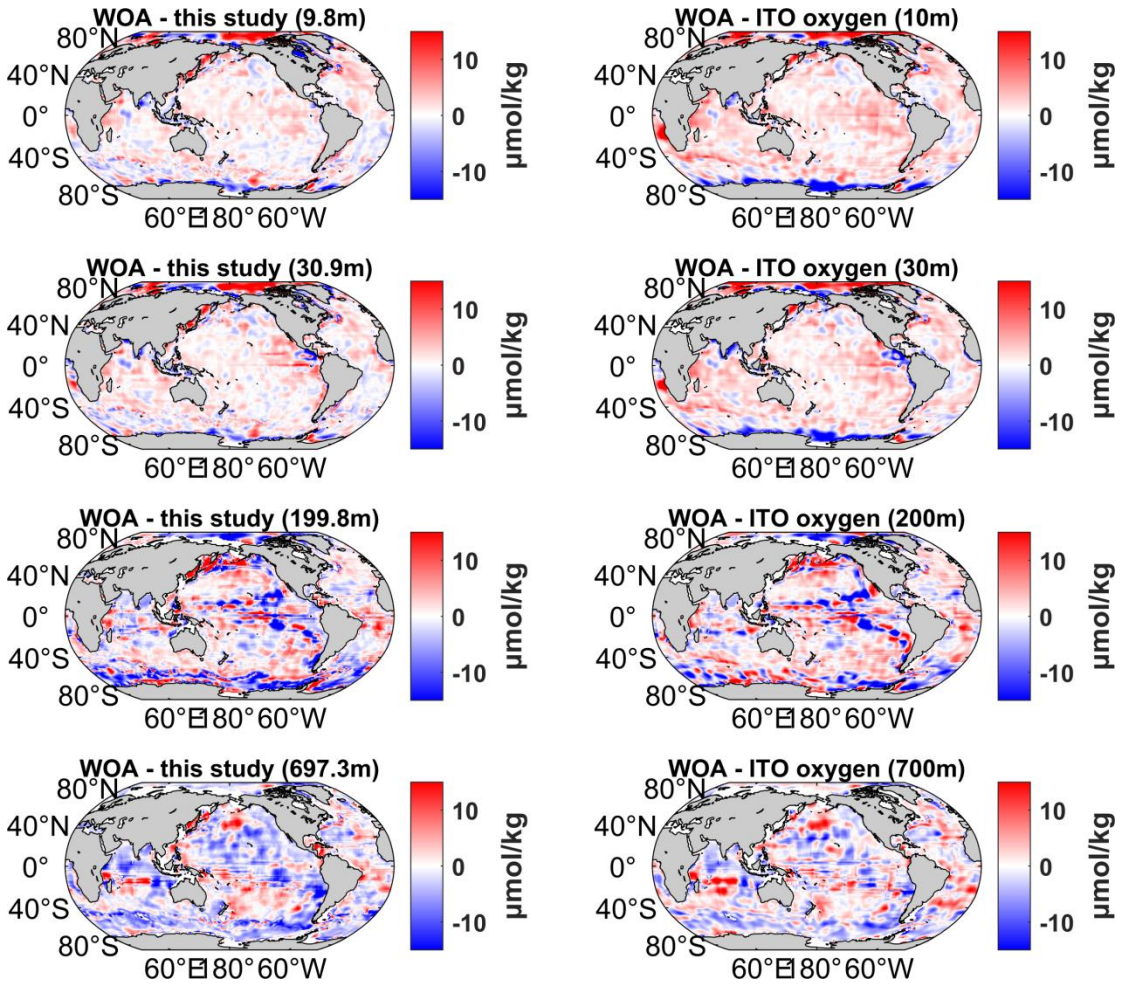

Figure R2. Spatial differences from WOA23 at four representative depths for this study and ITO Oxygen. Left panels show WOA23 minus this study at 9.8, 30.9, 199.8, and 697.3 m. Right panels show WOA23 minus ITO Oxygen at 10, 30, 200, and 700 m. Units are µ mol kg^{-1}.

Comment 2:

2. Spatial-Temporal Representativeness of the Validation Set

The authors state that for each profile in GLODAPv2, they searched the CTD and OSD records for matches within ±1° and the same month, excluding those that matched. This filtered the dataset down to 8,020 profiles. However, the manuscript lacks a spatial-temporal distribution map of this filtered validation set. It is critical to prove the coverage and representativeness of these remaining 8,020 profiles. Without a distribution map showing the cruise tracks or sampling locations, readers cannot determine whether the validation set represents a global oceanic assessment or if it is merely biased toward a few localized, data-rich sub-regions. Please provide maps and temporal histograms of the validation set.

Response 2:

We thank the reviewer for this suggestion. To clarify the coverage and representativeness of the filtered GLODAPv2 validation set, we have added a new figure showing both its spatial distribution and temporal histograms (Figure R3). The filtered dataset contains 8,020 unique profiles and spans all major ocean basins, indicating that the validation is not limited to a small number of localized regions. The yearly histogram shows that the profiles are distributed across multiple decades, although sampling becomes denser in the more recent period, while the monthly histogram shows coverage throughout the year with a clear peak in July.

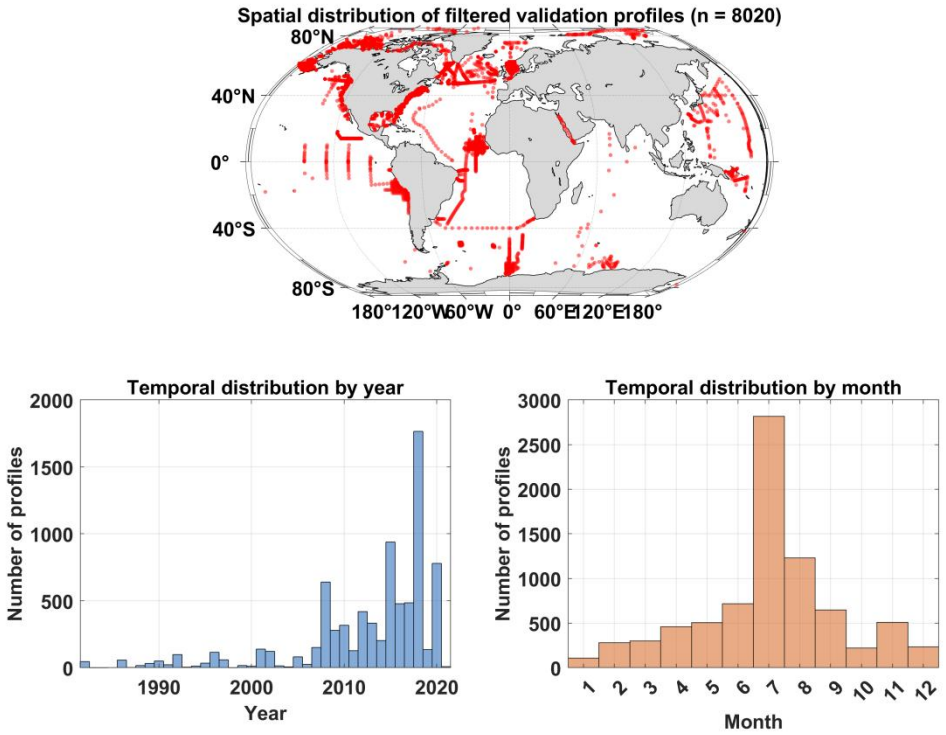

Figure R3. Spatial and temporal distribution of the filtered GLODAPv2 validation set. The upper panel shows the locations of the 8,020 unique validation profiles after excluding profiles matched with the CTD and OSD records used in model training. The lower panels show the temporal distribution of these profiles by year and by month.

Comment 3:

3. Potential Spatial Discontinuity in the Weight Allocation Strategy

While the dynamic weighting strategy is conceptually interesting, its current mathematical formulation may benefit from further justification. The transition mechanism between dynamic and static weights could potentially lead to spatial discontinuities. For instance, suppose grid cell A contains an observation, and the adjacent grid cell B does not. In cell A, the dynamic weight might heavily favor a specific model that perfectly fits the local observation; however, in cell B, the weight instantaneously reverts to the global average static weights (w_i) of the 6 models. This abrupt transition ("hard switch") between observed and unobserved regions might produce artificial gradients or step-changes at the boundaries, which may not fully align with the continuous nature of oceanographic variables. I recommend the authors discuss this potential limitation to ensure physical continuity.

Response 3:

Thank you for this comment. We carefully re-examined our original dynamic weighting strategy in light of your concern. Although we did not observe obvious spatial discontinuities in the final reconstructed dissolved oxygen field, we agree that the reviewer's concern is theoretically well founded. In the original formulation, the model weight at an observed grid cell was determined by the collocated local error, whereas at a neighboring unobserved grid cell the weight reverted directly to the global prior weight. This observation-dependent switching could indeed introduce an abrupt transition in the weight field. To address this issue, we revised the dynamic weighting scheme by introducing a spatially smoothed weighting framework that preserves the original local error-based weighting at observation-supported grid cells while allowing the influence of these locally constrained weights to extend continuously into neighboring regions. The revised formulation remains conceptually consistent with ensemble weighting based on model skill and with locally calibrated weighting strategies that use nearby observations to inform local adjustment (Raftery et al., 2005; Kleiber et al., 2011; Brunsdon et al., 1996).

First, the global prior weight of model i is still defined from its time-cross-validation RMSE ε_i as

$$\omega_i = \frac{\exp(-\beta\varepsilon_i)}{\sum_{j=1}^{M}\exp(-\beta\varepsilon_j)},$$

where M=6 is the number of base models and β is the prior-weight sensitivity parameter. At grid cells with valid observations, we retain the original local error-based weighting by first defining the local score of model i as

$$s_i(x) = \exp\left[-\alpha\left|p_i(x) - O(x)\right|\right],$$

where $p_i(x)$ is the prediction of model i, O(x) is the observation, and α controls the sensitivity of the local weighting to model error. These scores are then normalized across all M models to obtain the effective local weight at observation-supported grid cells:

$$l_i^{obs}(x) = \frac{s_i(x)}{\sum_{j=1}^{M} s_j(x)}.$$

To introduce spatial continuity, we then smooth these effective local weights using neighboring

observation-supported grid cells weighted by a Gaussian kernel, following the general idea of kernel-weighted local estimation for spatially varying relationships (Brunsdon et al., 1996):

$$K_{\mathrm{h}}(x, x_n) = \exp\left(-\frac{d_{xy}(x, x_n)^2}{2\sigma_{xy}^2} - \frac{d_z(x, x_n)^2}{2\sigma_z^2}\right),$$

$$\widetilde{l}_i(x) = \frac{\sum_{x_n \in N(x)} K_h(x, x_n) l_i^{obs}(x_n)}{\sum_{x_n \in N(x)} K_h(x, x_n)},$$

where x is the target grid cell, x_n denotes neighboring observation-supported grid cells and N(x) is the set of neighboring locations with valid effective local weights. In this way, the revised method retains the original locally constrained weighting at observed locations while producing a spatially continuous extension of these weights into neighboring areas.

To avoid another abrupt transition between observation-rich and observation-sparse regions, we further define an observation-support factor

$$S(x) = \sum_{x_n \in N(x)} K_h(x, x_n),$$

$$\rho(x) = \frac{S(x)}{S(x) + c},$$

where c is a shrinkage parameter. The final model weight is then written as

$$w_i(x) = \rho(x)\widetilde{l}_i(x) + \left[1 - \rho(x)\right]\omega_i,$$

Finally, the ensemble reconstruction is calculated as

$$\hat{O}(x) = \sum_{i=1}^{M} w_i(x) p_i(x).$$

Under this revised formulation, the influence of observations no longer changes instantaneously between observed and unobserved grid cells. Instead, the effective local weights derived at observation-supported grid cells are propagated continuously through the kernel function, and these smoothed local weights are gradually shrunk toward the global prior weights as observational support decreases. Therefore, the revised method preserves the strengths of the original framework, namely local observational constraint and global prior stability, while improving the spatial smoothness of the weight field.

Because RF generally receives relatively higher weights than the other five models in observation-supported regions, its spatial weight pattern provides a clearer illustration of the transition from the discontinuous structure in the original method to the more continuous structure in the revised method. As shown in Figure R4, we compared the global mean surface weight distribution of the RF model at 0.5 m obtained from the original and revised weighting strategies. The original method exhibits uneven spatial patterns and more abrupt transitions in the RF weight field, particularly in regions where observational coverage changes rapidly, which is consistent with the reviewer's concern regarding potential discontinuities induced by the hard switch. In contrast, the revised method produces a much smoother and more continuous global weight distribution. The sharp transitions present in the original formulation are reduced, and the influence of observation-rich regions extends more gradually into neighboring areas. This

comparison shows that the revised weighting framework reduces the potential spatial discontinuity in the original method. We have revised the manuscript accordingly by replacing the previous formulation with the new smoothed dynamic weighting scheme.

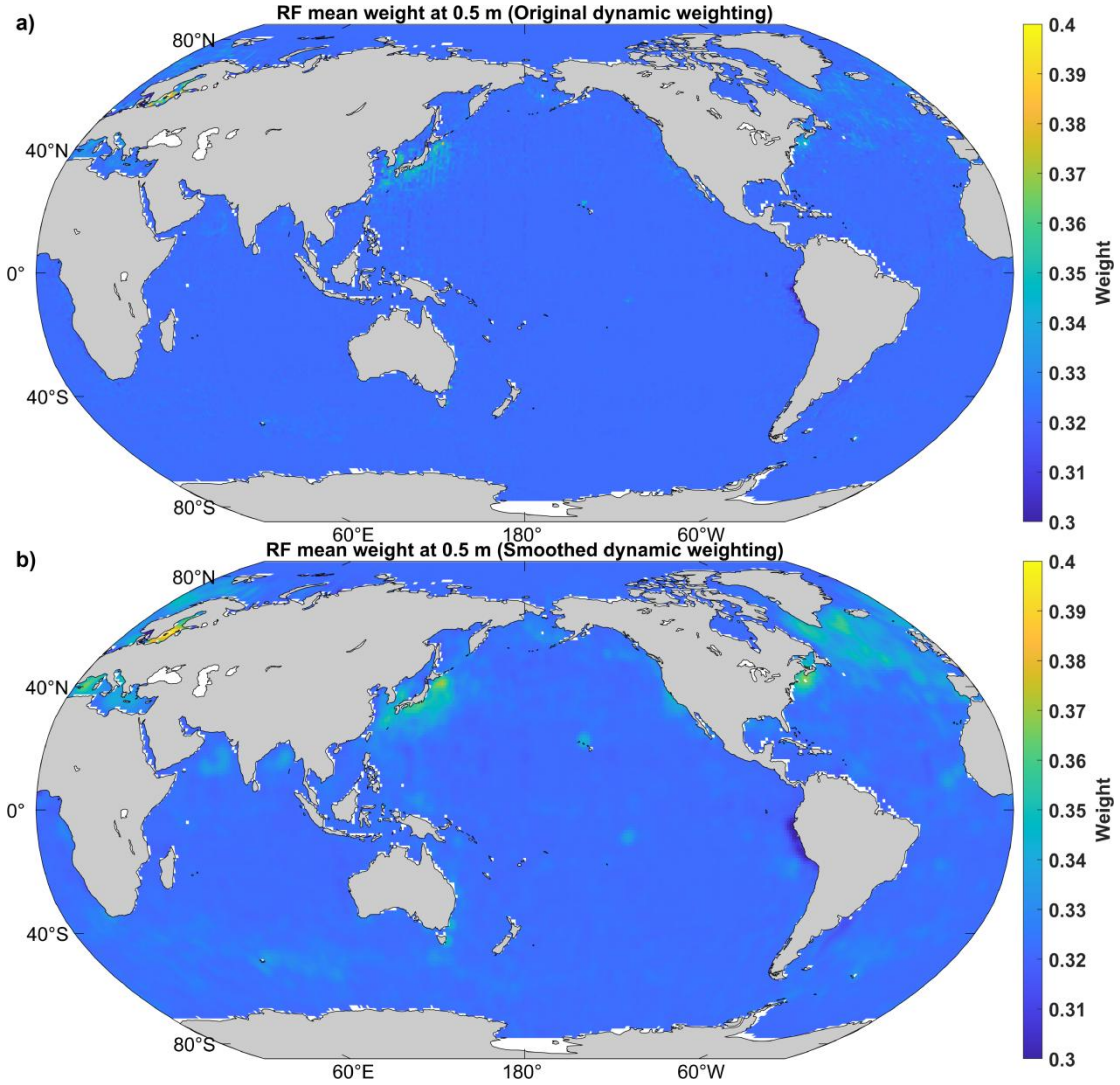

Figure R4. Global mean RF weight distribution at 0.5 m under the (a) original weighting scheme and (b) revised smoothed dynamic weighting scheme.

Although the weighting strategy was revised, the overall dissolved oxygen fields remain essentially unchanged. As shown in Figure R5, the global mean dissolved oxygen distributions produced by the original and revised methods are nearly identical at representative depths from the surface to the deep ocean (0.5, 199.8, 856.7, and 4093.2 m). No obvious basin-scale shifts are introduced by the revised method. This indicates that the new smoothed dynamic weighting mainly improves the spatial continuity of the weighting field, while maintaining the original reconstruction results and their large-scale oceanographic structure.

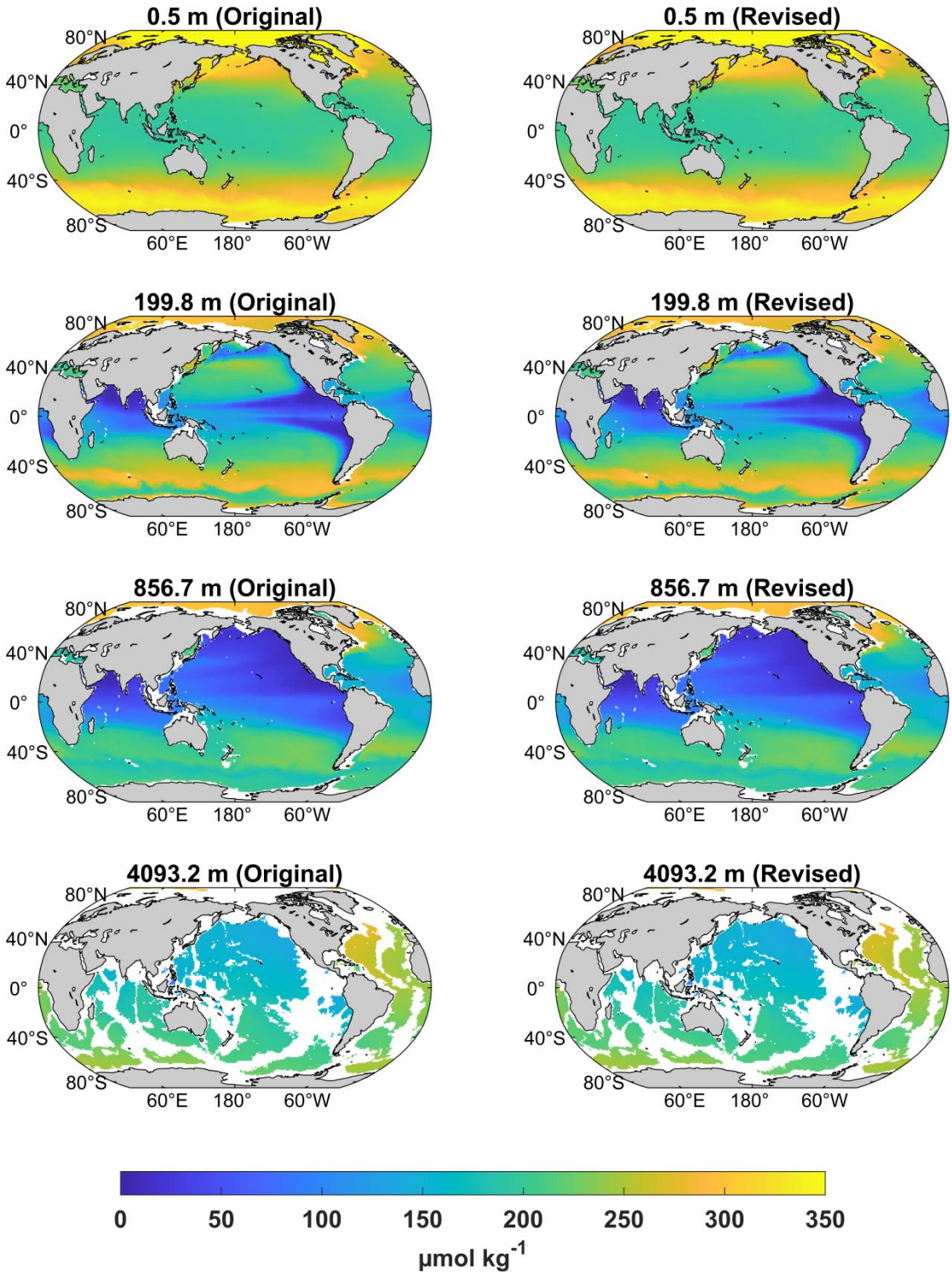

Figure R5. Global distributions of time-mean dissolved oxygen concentration at four representative depths, 0.5, 199.8, 856.7, and 4093.2 m, for the original reconstruction (left column) and the revised reconstruction (right column). Values were averaged over all months and years in the study period.

Comment 4:

4. Insufficient Assessment of Deep-Ocean Accuracy

A major selling point of this dataset is its extension to 5,902 m depth. However, the manuscript lacks a rigorous, depth-specific accuracy assessment for the deep ocean. Validating the entire water column collectively obscures potential biases in the bathypelagic zone. To substantiate the claims regarding deep-ocean reconstruction, I suggest conducting a direct comparison of your deep-ocean results with the DIVA-based dataset by Roach and Bindoff (2023). This comparison will help verify if your machine-learning ensemble correctly captures the subtle deep-water mass structures compared to variational analysis methods.

Reference: Roach, C. J., & Bindoff, N. L. (2023). Developing a New Oxygen Atlas of the World's Oceans Using Data Interpolating Variational Analysis. Journal of Atmospheric and Oceanic Technology.

Response 4:

We thank the reviewer for this suggestion. We agree that, because our product extends to 5902 m, a depth-specific assessment is needed to evaluate its performance in the deep ocean rather than relying only on whole-water-column statistics. To address this point, we added a direct comparison with the DIVA-based dataset of Roach and Bindoff (2023), which extends to 6800 m and provides an independent reference for deep-ocean structure. Figures R6 and R7 show the 1965–2022 mean vertical profiles from our reconstruction, ITO Oxygen, WOA23, and Roach and Bindoff (RB), with Figure R6 focusing on the upper 1000 m and Figure R7 extending the comparison to 6000 m. In the upper 50 m, our reconstruction and RB show very similar profiles, and both are close to WOA23. Between about 100 and 400 m, RB is slightly closer to WOA23, whereas from about 400 to 1000 m our profile approaches WOA23 and remains close to both reference products. From 1000 to 3500 m, the profiles of our reconstruction, RB, and WOA23 are nearly indistinguishable, indicating very similar large-scale deep-ocean structure in this depth range. Below 3500 m, the RB profile remains slightly closer to WOA23, but the differences among the profiles are still small.

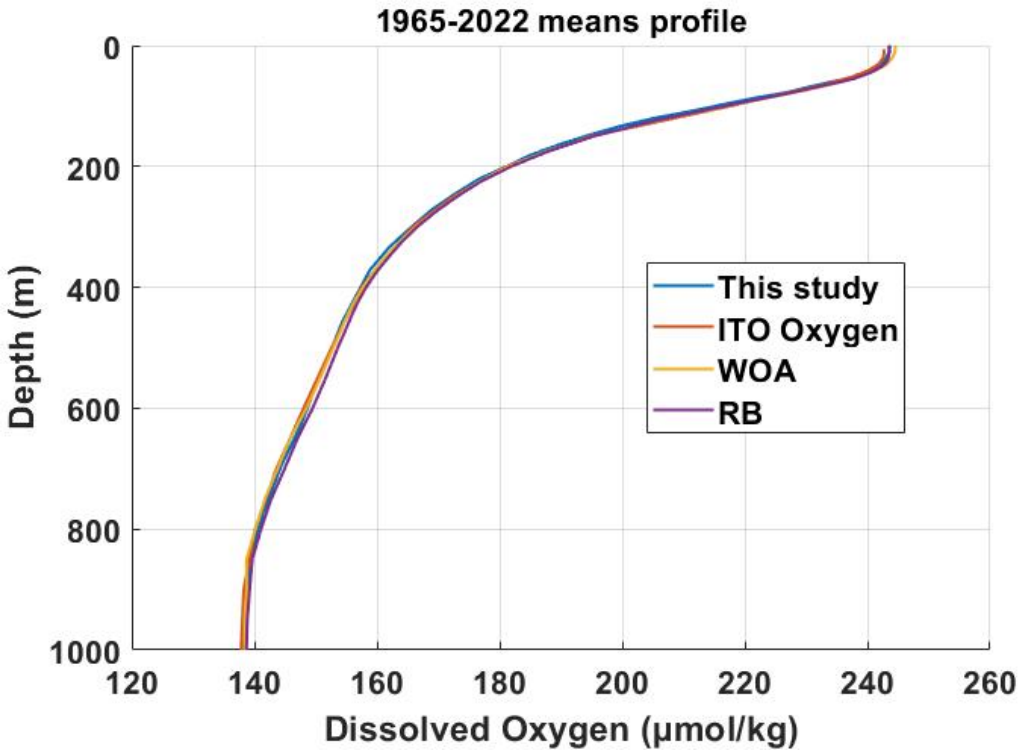

Figure R6. Global mean vertical profiles of different dissolved oxygen products (1965 – 2022). Solid lines show our reconstruction (blue), Roach & Bindoff's reconstruction (purple), ITO Oxygen (orange) and WOA23 climatology (yellow), plotted from the surface down to 1000 m.

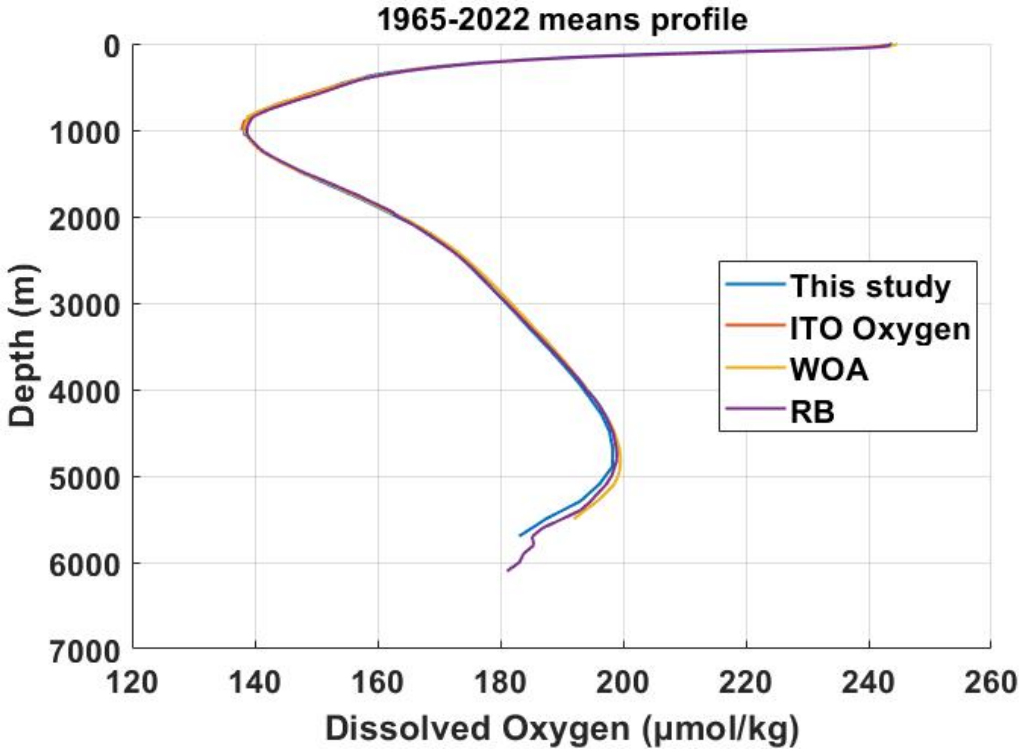

Figure R7. Global mean vertical profiles of different dissolved oxygen products (1965 – 2022). Solid lines show our reconstruction (blue), Roach & Bindoff's reconstruction (purple), ITO Oxygen (orange) and WOA23 climatology (yellow), plotted from the surface down to 6000 m.

We further added a spatial comparison with WOA23 at four representative depths to assess the deep-ocean performance of our reconstruction against the Roach and Bindoff product (Figure R8). At the surface layer around 10 m, our reconstruction is generally closer to WOA23 than RB, with smaller spatial differences over much of the open ocean. At around 700 m, both our reconstruction and RB show relatively large differences from WOA23. At around 1500 m, the differences decrease in both products, and at around 4000 m they decrease further, with both products showing generally smaller differences relative to WOA23. Overall, this comparison indicates that our product agrees more closely with WOA23 in the surface layer, while in the deep ocean both our reconstruction and RB show reduced differences below 1500 m.

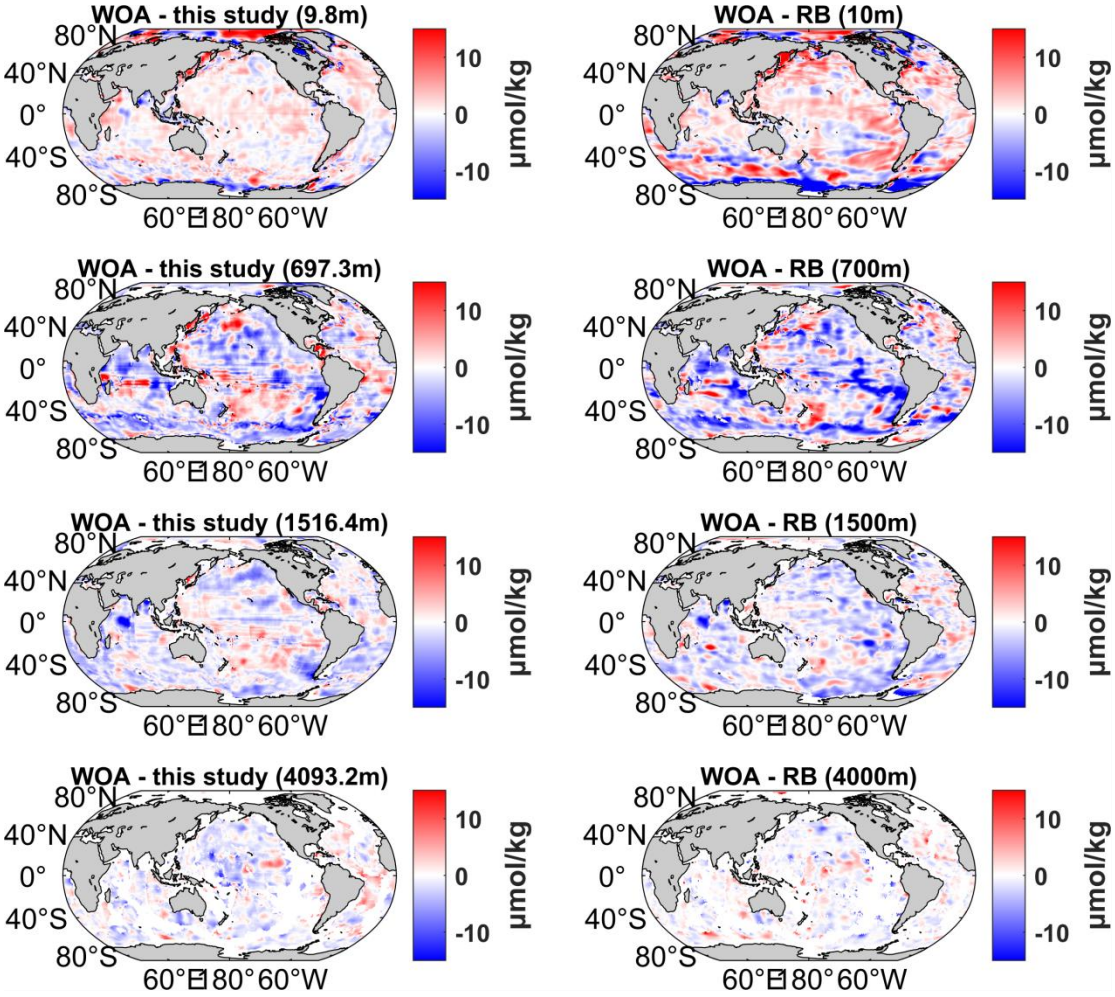

Figure R8. Spatial differences from WOA23 at four representative depths for this study and the Roach and Bindoff (RB) product. Left panels show WOA23 minus this study at 9.8, 697.3, 1516.4, and 4093.2 m. Right panels show WOA23 minus RB at 10, 700, 1500, and 4000 m. Units are µmol kg^{-1}.

Other comments:

Response 5:

Thank you for the technical corrections. We have implemented every suggested change and double-checked the manuscript. Your detailed remarks have improved the clarity and accuracy of the manuscript.

Reference

Brunsdon C, Fotheringham A S, Charlton M E. Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical analysis*, **1996**, 28(4): 281-298.

Garcia, Hernan E., et al. *World Ocean Atlas 2023, Volume 3: Dissolved Oxygen, Apparent Oxygen Utilization, Dissolved Oxygen Saturation and 30-year Climate Normal.* **2024**.

Ito T, Cervania A, Cross K, et al. Mapping dissolved oxygen concentrations by combining shipboard and Argo observations using machine learning algorithms. *Journal of Geophysical Research: Machine Learning and Computation*, **2024**, 1(3): e2024JH000272.

Kleiber W, Raftery A E, Baars J, et al. Locally calibrated probabilistic temperature forecasting using geostatistical model averaging and local Bayesian model averaging. *Monthly Weather Review*, **2011**, 139(8): 2630-2649.

Olsen A, Key R M, Van Heuven S, et al. The Global Ocean Data Analysis Project version 2 (GLODAPv2)–an internally consistent data product for the world ocean. *Earth System Science Data*, **2016**, 8(2): 297-323.

Raftery A E, Gneiting T, Balabdaoui F, et al. Using Bayesian model averaging to calibrate forecast ensembles. *Monthly weather review*, **2005**, 133(5): 1155-1174.

Roach, C. J., Bindoff, N. L. Developing a new oxygen atlas of the world's oceans using data interpolating variational analysis. *Journal of Atmospheric and Oceanic Technology*, **2023**, 40(11): 1475-1491.

Sharp J D, Fassbender A J, Carter B R, et al. GOBAI-O2: temporally and spatially resolved fields of ocean interior dissolved oxygen over nearly 2 decades. *Earth System Science Data*, **2023**, 15, 4481–4518, https://doi.org/10.5194/essd-15-4481-2023.